REMARKS

The accompany Continued Prosecution Application is being filed in response to the March 13, 2001 Advisory Action. On March 1, 2001, the applicants filed a response to a Final Action mailed on December 5, 2000. In the May 13, 2001 Advisory Action, the Examiner indicated that the request for reconsideration had been considered but the request did not place the application in condition for allowance because of the reasons set forth in the Continuation Sheet accompanying the May 13, 2001 Advisory Action.

Prior to examination of the above-identified Continued Prosecution Application, please enter the above amendments to the claims and consider the following remarks. Claims 1-35 are pending. Claims 1, 9, 17 and 25-27 have been amended. Claims 28-35 have been added. Reconsideration and allowance of the pending claims in view of the above amendments and the following remarks are respectfully requested.

Claims 1, 3, 5, 8, 9, 11, 13, 16-20 and 24-27 were rejected in the December 5, 2000 Final Action under 35 U.S.C. §102(e) as being anticipated by Kageyama (U.S. Patent No. 5,955,693). Claims 4, 14 and 22 were also rejected in the December 5, 2000 under 35 U.S.C. §103(a) as being unpatentable over Kageyama. These rejections are respectfully traversed.

Embodiments of the present invention are directed to voice converters that allow imitation of a professional singer to be performed by a karaoke player. In a preferred embodiment, the voice converter includes an extracting means that extracts a plurality of sinusoidal wave components from an input voice signal, where the sinusoidal wave components are spectral wave components of the input voice in the form of frequency value coordinates and/or amplitude value coordinates. For example, each sinusoidal component may be represented by a parameter set in the form of frequency value F and amplitude value A coordinates, such as (F_0, A_0), (F_1, A_1), (F_2, A_2), ...(F_n, A_n), where n is an integer (see Figs. 2 and 3). The frequency value coordinates in each of the parameter set represent the frequency of

8

the sinusoidal wave component corresponding to the parameter set. The amplitude value coordinates in each of the parameter sets represent the amplitude of the sinusoidal wave component corresponding to the parameter set. A modulating means is provided to modulate the frequency value coordinates and/or the amplitudes value coordinates, and thus the frequencies and/or amplitudes of the sinusoidal wave components, according to pitch information and/or amplitude information of a reference voice signal. The modulation is depicted in Fig. 6. The reference voice signal may, for example, represent a professional singer's voice signal that a karaoke singer is trying to imitate. After the modulation, a mixing means mixes the plurality of sinusoidal wave components to synthesize an output voice signal having a pitch different from that of the input voice signal and influenced by that of the reference voice signal.

In an illustrative example, as shown in Fig. 1 of the current application, a microphone 1 gathers a karaoke singer's voice and provides an input voice signal Sv. The input voice signal Sv is then analyzed by a Fast Fourier Transform (FFT) section 2, and the frequency spectrum thereof is detected (see page 8, lines 12-16 of the current application). The processing implemented by the FFT section 2 is carried out in prescribed frame units, so that a frequency spectrum is created successively for each frame (see Fig. 2 of the current application). A peak detecting section 3 detects peaks in the frequency spectrum of the input voice signal Sv. For example, sampling results of the plurality of sinusoidal wave components Fn, An, as illustrated in part (1) of Fig. 6 of the current application, are obtained. The frequency Fn and the amplitude An of each sinusoidal component are modified according to a pitch and volume of a model voice signal, as depicted in Fig. 6 of the current application. Accordingly, when the voice of the karaoke singer is outputted after the modification, the characteristics of the voice, the manner of singing, and the like are significantly influenced by the model voice signal.

The Kageyama reference discloses a karaoke apparatus capable of changing a live singing voice to a similar voice of an original singer of a karaoke song. However, Kageyama does not disclose an extracting means that extracts "a plurality of sinusoidal wave components from the input voice signal, <u>the sinusoidal wave components being spectral wave components of the input voice and in the form of at least frequency value coordinates</u>" and a modulating means for "<u>modulating frequency value coordinates of the sinusoidal wave components</u> of the input voice

signal according to the reference pitch information" representative of a pitch of a reference voice signal, as recited in amended claim 1 and similarly recited in amended claim 25. This limitation is hereinafter referred to as "the frequency value coordinate limitation". Likewise, Kageyama does not disclose an extracting means that extracts "a plurality of sinusoidal wave components from the input voice signal, the sinusoidal wave components being spectral wave components of the input voice and in the form of at least amplitude value coordinates" and a modulating means for "modulating amplitude value coordinates of the sinusoidal wave components of the input voice signal extracted from the input voice signal according to the reference amplitude information" representative of amplitudes of sinusoidal wave components contained in a reference voice signal, as is recited in amended claims 9 and similarly recited in amended claim 26. This limitation is hereinafter referred to as "the amplitude value coordinate limitation." Moreover, Kageyama also does not disclose limitations recited in claims 17 and 27. Claim 17 incorporates both the frequency value coordinate limitation and amplitude value coordinate limitation of claims 1 and 9, analyzing and modulating "a plurality of sinusoidal wave components contained in the input voice signal to derive a parameter set of an original frequency and an original amplitude, the sinusoidal wave components being spectral wave components of the input voice, the parameter set representing a corresponding sinusoidal wave component." Amended claim 27 contains similar recitations.

Kageyama discloses an apparatus and method of modifying a live singing voice (*i.e.*, input voice) to a voice (*i.e.*, model voice) similar to the original/model singer of the karaoke song. However, the voice modifying method is different from the present voice conversion using a set of sinusoidal components that represent spectral wave components of the input voice signal, with each sinusoidal component being, for example, in the form of a parameter set (Fn, An). The karaoke apparatus in the Kageyama reference uses phoneme data of a model singer to modify and approximate the voice of the live karaoke singer to that of the model singer. The phoneme data represents primary characteristics of the vowels contained in the model voice of the model singer, in terms of the waveform, envelope thereof, vibrato frequency, vibrato depth and

10

supplemental noise (see column 4, line 66 to column 5, line 2).[1] When the live singing voice is input in the karaoke apparatus in the Kageyama reference, a separating device separates the lead consonant component and the subsequent vowel component of the live singing voice. After the separation, an extracting device extracts the secondary characteristics of the subsequent vowel component, which may for example be the pitch of the separated subsequent vowel component. A substitutive vowel component is then created according to the primary characteristics of the vowels in the model voice (i.e., the phoneme data of model voice) and the secondary characteristics (e.g., the pitch of input voice). The substitutive vowel component, having the waveform of the model vowel and the pitch of the separated subsequent vowel component from the live singing voice, basically replaces the subsequent vowel component. Finally, the substitutive vowel component and is combined with the lead consonant component to synthesize an output singing voice.

In contrast, the present invention utilizes a voice conversion apparatus that <u>extracts a plurality of sinusoidal wave components from an input voice signal, the sinusoidal wave components being spectral wave components of the input voice and in the form of at least one of frequency value coordinates or amplitude value coordinates</u>." The set of sinusoidal wave components extracted from the input voice signal are <u>spectral wave components of the input voice signal</u>, and not a vowel portion of a syllable in the input voice signal or a digitized consonant portion and a digitized vowel portion of a syllable of the input voice, as disclosed in the Kageyama reference. The set of sinusoidal wave components may, for example, be in the form of frequency value and amplitude value coordinates (Fn, An), where n is an integer. Each set of (Fn, An) component represents a parameter set of the original frequency and the original amplitude of each sinusoidal wave component. All or some of the extracted sinusoidal wave components are then <u>modulated by reference pitch information of a reference voice signal and/or reference amplitude information of a reference voice signal</u>.

---

[1] Referring to Fig. 6A of the Kageyama reference, a phrase of lyric "A KA SHI YA NO" comprises five syllables "A", "KA", "SHI", "YA" and "NO", and the phoneme data are composed of extracted vowels "a", "a", "I", "a" and "o" from the five syllables.

In the karaoke apparatus of Kageyama, vowel components are extracted from the input voice and then replaced with a substitute vowel component having a waveform of a reference model vowel and a pitch or level of the input voice. The "pitch" mentioned in the Kageyama reference refers to the fundamental frequency of the digitized vowel portion of a syllable contained in the input singing voice signal. This is not the same as frequency value coordinate Fn of the present invention, which represents frequency of the sinusoidal wave component in the frequency spectral wave components F1-Fn of an input voice signal. The "level" mentioned in the Kageyama reference refers to the volume or envelop of the digitized vowel portion of a syllable contained in the input singing voice signal. This is not the same as amplitude value coordinate An of the present invention, which represents amplitude of the sinusoidal wave component in the amplitude spectral wave components A1-An of an input voice signal.

The karaoke apparatus of Kageyama does not disclose, teach or suggest extracting a set of sinusoidal wave components, the sinusoidal wave components being spectral wave components of the input voice and in the form of at least one of frequency value coordinates or amplitude value coordinates, and modulating these components. Likewise, the karaoke apparatus of Kageyama does not disclose, teach or suggest an analyzer device that analyzes a plurality of sinusoidal wave components contained in the input voice signal to derive a parameter set of an original frequency and an original amplitude, with the sinusoidal wave components being spectral wave components of the input voice and the parameter set representing a corresponding sinusoidal wave component. The karaoke apparatus of Kageyama also does not disclose, teach or suggest a modulator device that modulates the parameter set of the sinusoidal wave components according to reference information. Therefore, it is respectfully submitted that claims 1, 9, 17 and 25-27 distinguish over the Kageyama reference. Because each dependent claim incorporates all the limitations of its base claim(s), claims depending from 1, 9 and 17 also distinguish over the Kageyama reference. Claims 3, 5 and 8 depend directly or indirectly from claim 1. Claims 11, 13 and 16 depend directly or indirectly from claim 9. Claims 18-20 and 24 depend directly or indirectly from claim 17. The rejection of claims 1, 3, 5, 8, 9, 11, 13, 16-20 and 24-27 under 35 U.S.C. §102(e) and claims 4, 14 and 22 under 35 U.S.C. §103(a) should therefore be withdrawn.

12

Claims 2, 6, 10, 12 and 21 were also rejected in the December 5, 2000 Final Action under 35 U.S.C. §103(a) as being unpatentable over Kageyama in view of Matsumoto '303 (U.S. Patent No. 5,847,303). Claims 7, 15 and 23 were rejected under 35 U.S.C. §103(a) as being unpatentable over Kageyama in view of Matsumoto '907 (U.S. Patent No. 5,963,907). These rejections are respectfully traversed.

The claimed features of the present invention are not realized even if the teachings of the Matsumoto '303 reference or Matsumoto '907 reference are incorporated into Kageyama. Matsumoto '303 is directed to a voice processing apparatus that modulates an input voice signal into an output voice signal according to a set of parameters. Matsumoto '303 discloses a voice change parameter table of filter coefficients to control spectrum shape of varying pitch ranges for the purpose of providing more realistic sounding conversion between male and female voices (see Figs. 9 and 10; column 11, lines 3-26 of the Matsumoto '303 reference). An audio signal processor within the voice processing apparatus is configured by a parameter set to process the audio signal by modifying the frequency spectrum of the input voice. However, Matsumoto '303 does not disclose the inventive features of the present invention in extracting a plurality of sinusoidal wave components from an input voice signal representing frequency spectral wave components of the input voice signal, the sinusoidal wave components including frequency value coordinates of the sinusoidal wave components and modulating frequency value coordinates of the sinusoidal wave components according to pitch information representative of a pitch of a reference voice signal, as is recited in claim 1. Likewise, Matsumoto '303 does not disclose extracting a plurality of sinusoidal wave components from the input voice signal representing amplitude spectral wave components of the input voice signal, the sinusoidal wave components including amplitude value coordinates of the sinusoidal wave components and modulating amplitude value coordinate of the sinusoidal wave component extracted from the input voice signal according to the amplitude information representative of amplitudes of sinusoidal wave components contained in a reference voice signal, as is recited in claim 9. Claim 17 incorporates the above limitations of claims 1 and 9; therefore, it also distinguishes over the Matsumoto '303 reference.

Matsumoto '907 is directed to a voice converter that provides pitch and formant shifting of an input voice signal. Referring to Fig. 2 of the Matsumoto '907 reference, an audio filter 325 extracts the volume level of the input voice signal, and outputs the extracted volume level as first volume data V1. A second audio filter 326 extracts the volume level of an output voice signal, and outputs the extracted volume level as second volume data V2. A difference judging circuit 322 compares the first and second volume data V1 and V2 with each other, and determines a volume gain G and a distorting factor D which is supplied to a distortion circuit 321. When the volume of the output voice after conversion is smaller than that of the input voice, the volume gain G is increased. In contrast, the subject matter of claims 7, 15 and 23 in the present invention is to change the volume of an input singing voice in matching with the variation of the volume of the voice of a model singer. This allows the volume of an output voice signal to emulate the volume variation of the reference voice signal of the model singer. Such feature is not disclosed, taught or suggested by Matsumoto '907. Additionally, Matsumoto '907 does not disclose the inventive features of the present invention in extracting a plurality of sinusoidal wave components from an input voice signal <u>representing frequency spectral wave components of the input voice signal</u>, the sinusoidal wave components including frequency value coordinates of the sinusoidal wave components and modulating frequency value coordinates of the sinusoidal wave components according to pitch information representative of a pitch of a reference voice signal, as is recited in claim 1. Likewise, Matsumoto '907 does not disclose extracting a plurality of sinusoidal wave components from the input voice signal <u>representing amplitude spectral wave components of the input voice signal</u>, the sinusoidal wave components including amplitude value coordinates of the sinusoidal wave components and modulating amplitude value coordinates of the sinusoidal wave component extracted from the input voice signal according to the amplitude information representative of amplitudes of sinusoidal wave components contained in a reference voice signal, as is recited in claim 9. Claim 17 incorporates the above limitations of claims 1 and 9; therefore, it distinguishes over the Matsumoto '907 reference.

Applicant believes that the differences between Kageyama, Matsumoto '303, Matsumoto '907 and the present invention are clear in claims 1, 9 and 17, which set forth voice conversion and synthesizing apparatuses that utilize a plurality of sinusoidal wave components according to

14

embodiments of the present invention. Therefore, claims 1, 9 and 17 distinguish over the Kageyama, Matsumoto '303 and Matsumoto '907 references. Claims depending directly or indirectly from claims 1, 9 and 17, such as claims 2, 6, 7, 10, 12, 15, 21 and 23, also distinguish over the above references. Applicant further believes that the differences between Kageyama, Matsumoto '907 and the present invention are clear in claims 7, 15 and 23, which set forth apparatus that emulate volume variation of a model singer according to embodiments of the present invention. Therefore, the rejection of claims 2, 6, 7, 10, 12, 15, 21 and 23 under 35 U.S.C. § 103(a) should be withdrawn.
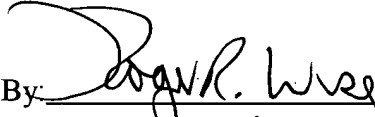
Claims 28-35 have been added by this preliminary amendment to further define the invention disclosed in the specification.

In view of the foregoing, it is respectfully submitted that the application and the claims are in condition for allowance.

If for any reason the Examiner finds the application other than in condition for allowance, the Examiner is invited to call the undersigned attorney at (213) 488-7100 to discuss the steps necessary for placing the application in condition for allowance, should the Examiner believe a telephone interview would advance prosecution of the application.

Respectfully submitted,

Dated: June ⟋, 2001   By: _____
                                  Roger R. Wise
                                  Reg. No. 31,204

Pillsbury Madison & Sutro LLP
725 South Figueroa Street, Suite 2800
Los Angeles, CA 90017
Telephone (213) 488-7100
Facsimile (213) 629-1033